

Zur Auswertung von Studienverläufen mit Process-Mining-Techniken

**Rüdiger Buck-Emden
Franz-Dominik Dahmann**

Publisher: Dean Prof. Dr. Wolfgang Heiden

**Hochschule Bonn-Rhein-Sieg – University of Applied Sciences,
Department of Computer Science**

Sankt Augustin, Germany

October 2017

Technical Report 07-2017



**Hochschule
Bonn-Rhein-Sieg**
University of Applied Sciences

ISSN 1869-5272

ISBN 978-3-96043-055-1

Copyright © 2017, by the author(s). All rights reserved. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

Das Urheberrecht des Autors bzw. der Autoren ist unveräußerlich. Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Das Werk kann innerhalb der engen Grenzen des Urheberrechtsgesetzes (UrhG), *German copyright law*, genutzt werden. Jede weitergehende Nutzung regelt obiger englischsprachiger Copyright-Vermerk. Die Nutzung des Werkes außerhalb des UrhG und des obigen Copyright-Vermerks ist unzulässig und strafbar.

Digital Object Identifier **doi:10.18418/978-3-96043-055-1**
DOI-Resolver **<http://dx.doi.org/>**

Zusammenfassung Studienverläufe von Studenten weichen nicht selten vom offiziell geplanten Curriculum ab. Für eine den Studienerfolg verbessernde Planung und Weiterentwicklung von Studiengängen und Curricula fehlen den Verantwortlichen häufig Erkenntnisse über tatsächliche sowie typischerweise erfolgreiche und weniger erfolgreiche Studienverlaufsmuster. Process-Mining-Techniken können helfen, mehr Transparenz bei der Auswertung von Studienverläufen zu schaffen und so die Erkennung typischer Studienverlaufsmuster, die Überprüfung der Übereinstimmung der konkreten Studienverläufe mit dem vorgegebenen Curriculum sowie eine zielgerechte Verbesserung des Curriculums zu unterstützen.

Abstract Students not always follow the curriculum officially planned for their course of studies. In order to optimize study programs and curricula according to student's needs and expected results, officials need detailed insights on how successful and less successful study patterns look like. Here process mining comes into play. It can help to understand typical study patterns, supports conformance checks between detected patterns and officially planned curricula, and provides valuable insights needed for curriculum planning and optimization.

Schlüsselwörter Process Mining, Studienverlauf, Curriculum, Analyse, Bubble-Chart, Fuzzy Miner, Inductive Visual Miner, RapidMiner, ProM, Disco

1 Einführung

Studienverläufe von Studenten an Hochschulen können mehr oder weniger erfolgreichen sein und weichen häufig und aus unterschiedlichen Gründen vom offiziell geplanten Curriculum ab. Für einen den Studienerfolg verbessernden Planung und Weiterentwicklung von Studiengängen und Curricula fehlen den Verantwortlichen (Studiendekane, Curriculums-Planer etc.) nicht selten Erkenntnisse über typischerweise erfolgreiche bzw. weniger erfolgreiche Studienverlaufsmuster. Die vorliegende Arbeit adressiert diesen Sachverhalt und untersucht, wie Process-Mining-Techniken (van der Aalst 2016, Rischawe und Buck-Emden 2015) mehr Transparenz bei der Analyse von Studienverläufen schaffen können. Diese Untersuchung soll verantwortlichen Studiendekanen und Curriculums-Planern Werkzeuge an die Hand geben, um z.B. Antworten auf die folgenden Fragen zu finden

- Was unterscheidet die Studienverläufe erfolgreicher Studenten von weniger erfolgreichen?
- Welche Lehrveranstaltungen werden von erfolgreichen bzw. weniger erfolgreichen Studenten wann im Studienverlauf belegt bzw. gemieden?
- An welchen Lehrveranstaltungen scheitern Studenten häufig?
- Welche Lehrveranstaltungen bringen Studenten letztendlich dazu, ihr Studium abzuberechnen?

Process Mining steht als Oberbegriff für Verfahren zur Analyse von Geschäftsprozessen auf Basis von digitalen Spuren (Log-Daten), die diese Geschäftsprozesse bei ihrer Ausführung im System hinterlassen haben. Die Analyseziele von Process Mining lassen sich in drei Kategorien zerlegen (van der Aalst 2016):

1. Entdecken (Discovery): Welches Prozessmodell liegt den beobachteten Log-Daten zugrunde?
2. Prüfen (Conformance): Entsprechen die beobachteten Abläufe einem vorgegebenen Prozessmodell?
3. Verbessern (Enhancement): Wie kann ein bestehendes Prozessmodell auf Basis der beobachteten Log-Daten verbessert werden?

Studienverläufe lassen sich nun ebenfalls als Geschäftsprozesse interpretieren, wobei jede relevante Aktivität (z.B. eine Prüfung) Protokollinformationen (z.B. beim Prüfungsamt) hinterlässt, die mithilfe von Process-Mining-Techniken analysiert werden können. Die Analyseziele entsprechen dabei den drei zuvor genannten: Entdecken der tatsächlichen Studienverlaufsmuster, Prüfen deren Übereinstimmung mit dem vorgegebenen Curriculum und Verbessern des Curriculums im Sinne einer gewünschten Zielfunktion (z.B. Durchschnittsnote, Studiendauer, Durchsatz).

Im Folgenden geben wir zunächst einen kurzen Überblick zum Stand der Forschung, beschreiben dann für uns relevante Werkzeuge und Verfahren des Process Minings und stellen abschließend ein darauf basierendes Vorgehensmodell zur Analyse von Studienverläufen vor.

2 Stand der Forschung

Process Mining hat seit dem wegweisenden Artikel von van der Aalst und Weijters aus dem Jahre 2003 (van der Aalst und Weijters 2003) kontinuierlich an Aufmerksamkeit gewonnen. Eine umfassende Einführung in die heutigen Möglichkeiten von Process Mining gibt van der Aalst in seinem aktuellen Buch zum Thema (van der Aalst 2016). Praktische Anwendungsszenarien von Process Mining sind vielfältig, werden aber schwerpunktmäßig im Kontext betrieblicher Abläufe diskutiert. Einige wenige Ansätze zur Nutzung von Process-Mining-Techniken für die Analyse von Studienverläufen an Hochschulen existieren, z.B. von Emonds und Scott 2015 und von Mukala et al. 2015. Unserem Vorschlag am nächsten kommen Trčka et al. 2010, allerdings mit einem weniger umfassenden Ansatz¹. Unser allgemeines, pragmatisches Vorgehensmodell wird im Folgenden vorgestellt.

¹ Trčka et al. 2010 beschränken sich auf die Bubble-Chart-Analyse für die Visualisierung von Prüfungsergebniskategorien auf Ebene einzelner Studenten und auf Fuzzy Mining zur Untersuchung der Popularität einzelner Lehrveranstaltungen.

3 Werkzeuge

Im Rahmen des in Kapitel 5 vorgestellten Vorgehensmodells wurden die folgenden Werkzeuge eingesetzt: Excel (Microsoft), Disco (Fluxicon), ProM (TU Eindhoven) und RapidMiner. Bis auf Microsofts Excel stehen diese Tools für Forschungszwecke kostenfrei zur Verfügung.

3.1 Excel

Microsoft Excel wurde für die Vorbereitung der zu analysierenden Studienverlaufsdaten genutzt. Passend aufbereitete Excel-Dateien können von vielen Analyse-Werkzeugen direkt als Eingabedaten verarbeitet werden. Im Rahmen unserer Untersuchung erforderte lediglich ProM eine Konvertierung in das eXtensible Event Stream(XES)-Format (Process Mining Group 2016)

3.2 RapidMiner

Der ursprünglich aus einem Forschungsprojekt an der TU Dortmund hervorgegangene, Java-basierte RapidMiner (Mierswa 2009) wird heute von der gleichnamigen Firma als Werkzeug für maschinelles Lernen und Data-Mining angeboten. Wir setzen im Rahmen dieser Untersuchung den RapidMiner Version 7.2 für die Bubble-Chart-Analyse ein (siehe Kapitel 4.1).

3.3 ProM

ProM ist ein an der TU Eindhoven entwickeltes, auf der Programmiersprache Java basierendes Process-Mining-Framework (van Dongen et al. 2005). Durch Verwendung von Plug-ins können verschiedenste Process-Mining-Verfahren angewendet werden. Im Rahmen dieser Untersuchung nutzen wir ProM Version 6.5 speziell mit dem Inductive Visual Miner (IvM, siehe Kapitel 4.3).

3.4 Disco

Disco ist eine komfortable, von der Firma Fluxicon entwickelte Process-Mining-Lösung (Günther and Rozinat 2012). Die Software stellt eine Weiterentwicklung des von ProM bekannten Fuzzy-Mining-Ansatzes mit einer verbesserten Usability und ansprechender Visualisierung der Analyseergebnisse dar. Disco ist in der Lage, Excel-Dateien direkt einzulesen und z.B. auch in XES zu exportieren (z.B. für die Weiterverarbeitung mit ProM). Von uns wurde Disco Version 1.9 eingesetzt.

4 Verfahren

Für die Analyse von Studienverläufen haben sich im Rahmen unserer Untersuchung besonders die nachfolgend skizzierten Verfahren Bubble-Chart-Analyse, Fuzzy Mining und Inductive Visual Mining als hilfreich erwiesen.

Bei den im Folgenden betrachteten Studienverläufen gehen wir beispielhaft von einem fiktiven, sechssemestrigen Studiengang aus, also von sechs Lehrplansemestern. Studenten benötigen sechs oder mehr Fachsemester, um das Studium abzuschließen (sofern sie das Studium nicht vorher abbrechen).

Studienverläufe interpretieren wir als Folgen von Prüfungen. Studenten können sich, sofern sie die Voraussetzungen erfüllen, in jedem Fachsemester zu Prüfungen der verschiedenen Lehrplansemester anmelden. Sie können diese Prüfungen entweder bestehen oder nicht bestehen.

4.1 Bubble-Chart-Analyse

Bubble-Charts (Blasendiagramme) sind eine Variante der Scatter-Charts (Streudiagramme), die eine zweidimensionale Darstellung mehrdimensionaler Daten ermöglichen.

Im Rahmen der Studienverlaufsanalyse ist die Bubble-Chart-Analyse besonders für eine überblicksartige Betrachtung des Studiums als Ganzes geeignet. Die mit dem RapidMiner durchgeführte Bubble-Chart-Analyse liefert in unserem Anwendungsfall eine vierdimensionale Koordinatendarstellung: X- und Y-Achse sowie Größe und Farbe der in das Koordinatenkreuz eingetragenen Objekte. Die X-Achse bezeichnet den Zeitverlauf (konkret: den Verlauf der ersten sechs Fachsemester), die Y-Achse die Prüfungen zu den einzelnen Lehrveranstaltungen. Die Objekte im Koordinatensystem repräsentieren einzelne Prüfungen. Ihre Farbe kennzeichnet die Durchfallquote und ihre Größe die Anzahl der teilnehmenden Studenten. Im Idealfall ergibt sich daraus das in Abb. 1 beispielhaft gezeigte Muster mit über sechs Lehrplansemester verteilten Lehrveranstaltungen V_1_1 bis V_6_4.

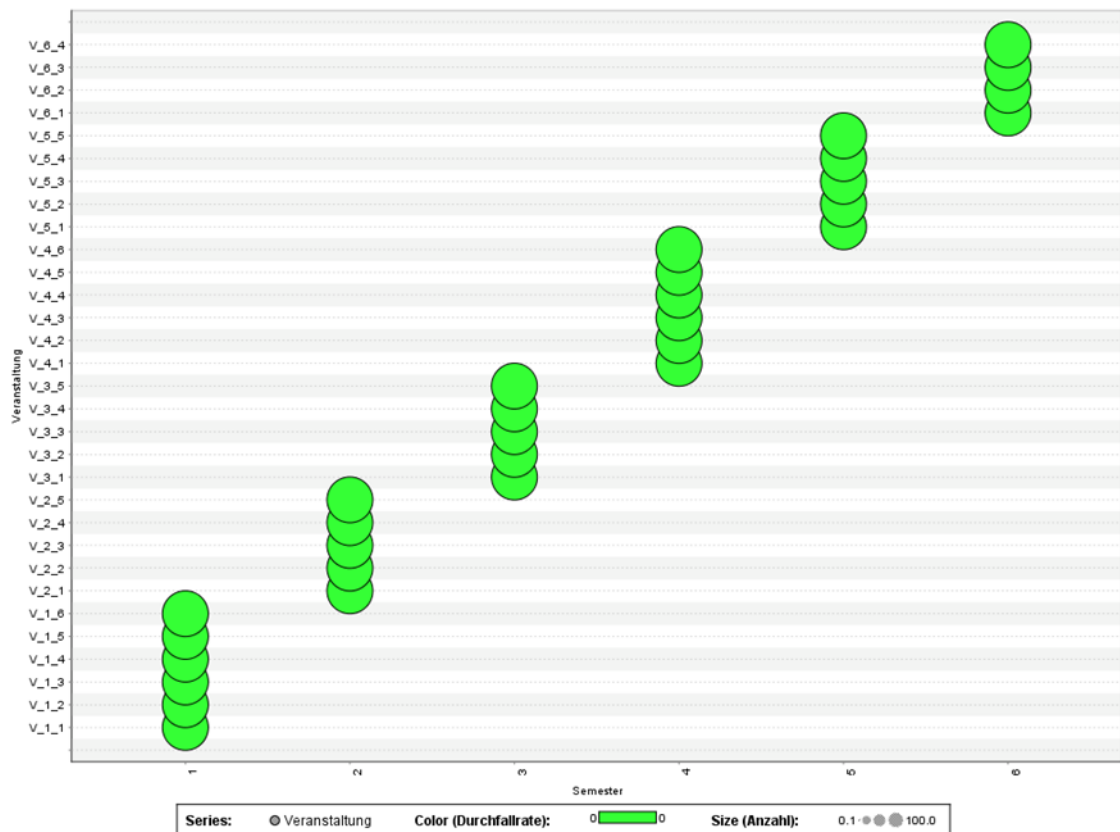


Abbildung 1: Beispielhafter Idealfall der Bubble-Chart-Analyse für Studienverläufe

Im Idealfall ist die volle Kapazität der Studienplätze zu Beginn des Studiums mit Studenten ausgelastet und alle Studenten erbringen alle Prüfungsleistungen in Regelzeit und in dem vom Curriculum vorgesehenen Lehrplansemester, ohne in einzelnen Prüfungen durchzufallen. Dadurch ergibt sich das in Abb. 1 dargestellte ideale Stufenmuster. In der Realität zeigen sich allerdings Abweichung vom idealen Verlaufsmuster, so dass nicht mehr alle Blasen die gleiche Farbe oder Größe haben und dass zu einzelnen Prüfung mehrere Blasen in verschiedenen Semestern auftreten. Dieser Fall tritt z.B. auf, wenn Studenten in einem Semester eine Prüfung zu einer Veranstaltung nicht bestanden haben und diese Prüfung in einem der Folgesemester nachholen. Je größer die tatsächliche Abweichung von dem in der Abbildung dargestellten Muster ist, desto mehr Studenten ist es nicht gelungen, sich an das vorgegebene Regel-Curriculum zu halten.

4.2 Fuzzy Mining

Fuzzy Mining ist mit ProM und Disco verfügbar. Auf Basis geeigneter Signifikanz- und Korrelationsmetriken für die graphischen Modellelemente unterstützt der Fuzzy Miner Benutzer bei der interaktiven Modellerzeugung auf unterschiedlichen Abstraktionsebenen (Günther und van der Aalst, 2007). Das Verfahren kann auch bei komplexen Geschäftsprozessen, die viele, miteinander vernetzte Aktivitäten umfassen, übersichtliche Prozessmodelle erzeugen, hat aber Limitationen bei der Darstellung paralleler Aktivitäten.

Zur Visualisierung der erzeugten Modelle verwendet der Fuzzy-Mining-Ansatz Rechtecke mit abgerundeten Ecken für Aktivitäten und gerichtete Kanten zwischen Aktivitäten als grafische Kennzeichnung der Reihenfolge, in unserem Fall der Studien- bzw. Prüfungsverläufe.

Die vom Benutzer interaktiv durchzuführende Fuzzy-Mining-Analyse besteht aus drei Schritten. Der erste Schritt ist die Konfliktlösung (wenn sich zwischen zwei Aktivitäten A und B Kanten $A \rightarrow B$ und $B \rightarrow A$ befinden, siehe Günther und van der Aalst 2007), der zweite das Kanten-Filtern und der dritte Schritt ist die Knotenaggregation und -abstraktion. In den ersten beiden Schritten werden Beziehungen zwischen Aktivitäten, die auf dem eingestellten Abstraktionsniveau keine Relevanz haben, entfernt. Im dritten Schritt werden dann je nach gewünschtem Abstraktionsgrad Aktivitäten zu Clustern zusammengefasst oder ebenfalls entfernt.

Fuzzy Mining hat sich im Rahmen der Studienverlaufsanalyse als besonders geeignet für die Detailbetrachtung einzelner Lehrveranstaltungen (oder genauer: der zugehörigen Prüfungen) erwiesen. Abb. 2 zeigt beispielsweise die Lehrveranstaltung V_1_6 mit den von einem Studentenjahrgang (Kohorte) im ersten, zweiten und dritten Versuch bestanden (BE) bzw. nicht bestanden (NB) Prüfungen. Wiederholungsprüfungen konnten in unterschiedlichen Fachsemestern stattfinden (siehe die Erläuterungen zur Bubble-Chart-Analyse in Kapitel 4.1) und schließen auch durchgeführte Verbesserungsversuche ein (in Abb. 2 gab es z.B. vier Verbesserungsversuche (2+2), von denen zwei nicht bestanden wurden).

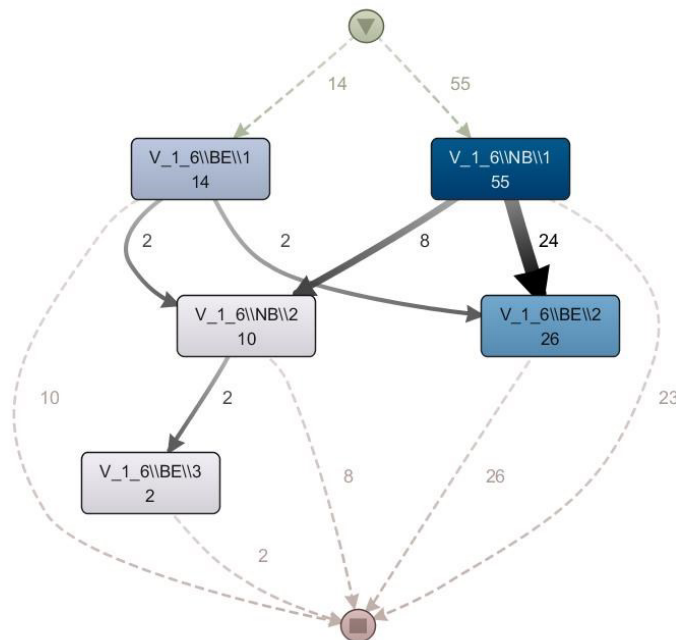


Abbildung 2: Fuzzy Mining einer Kohorte zu Lehrveranstaltung/Prüfung V_1_6

4.3 Inductive Visual Mining

Inductive Visual Mining (IvM) unterstützt ebenfalls die interaktive Erzeugung von Prozessmodellen, kann aber im Unterschied zu Fuzzy Mining auch parallel ablaufende Aktivitäten darstellen. Durch Filterung von als nicht relevant erachteten Aktivitäten und Kanten sind unterschiedliche Abstraktionsniveaus möglich. Die mit Inductive Visual Mining erstellten Prozessmodelle sind ausführbar und auf Ebene einzelner Prozessinstanzen visualisierbar (Leemans et al. 2013, Leemans et al. 2014).

Wie bei Fuzzy Mining stellen beim Inductive Visual Mining Rechtecke mit abgerundeten Ecken Aktivitäten und gerichtete Kanten die Reihenfolge, in der Aktivität ausgeführt wurden, dar. Ferner gibt es für Kanten UND-Splits und -Joins (Raute mit

Plus-Zeichen) sowie ODER-Splits und -Joins (kleiner Kreis) sowie gelbe Kreisobjekte für einzelne Prozessinstanzen.

Inductive Visual Mining ermöglicht eine detaillierte Auswertung von Folgen von Prüfungen, also von Prüfungsverläufen, auf Ebene der einzelnen Studenten. Abb. 3 zeigt beispielsweise für eine Kohorte mit 72 Studenten, wie viele von ihnen bis zum sechsten Fachsemester die einzelnen Prüfungen des zweiten. Lehrplansemesters absolviert haben, wie viele sich für einzelne Prüfungen nicht angemeldet haben (obwohl diese laut Studienplan vorgesehen waren, rote gestrichelte Linie²), wie viele Prüfungswiederholungen es gab (Rückpfeile) und wie viele Studenten einen größeren Block, bestehend aus mehreren Prüfungen, komplett umgangen haben³.

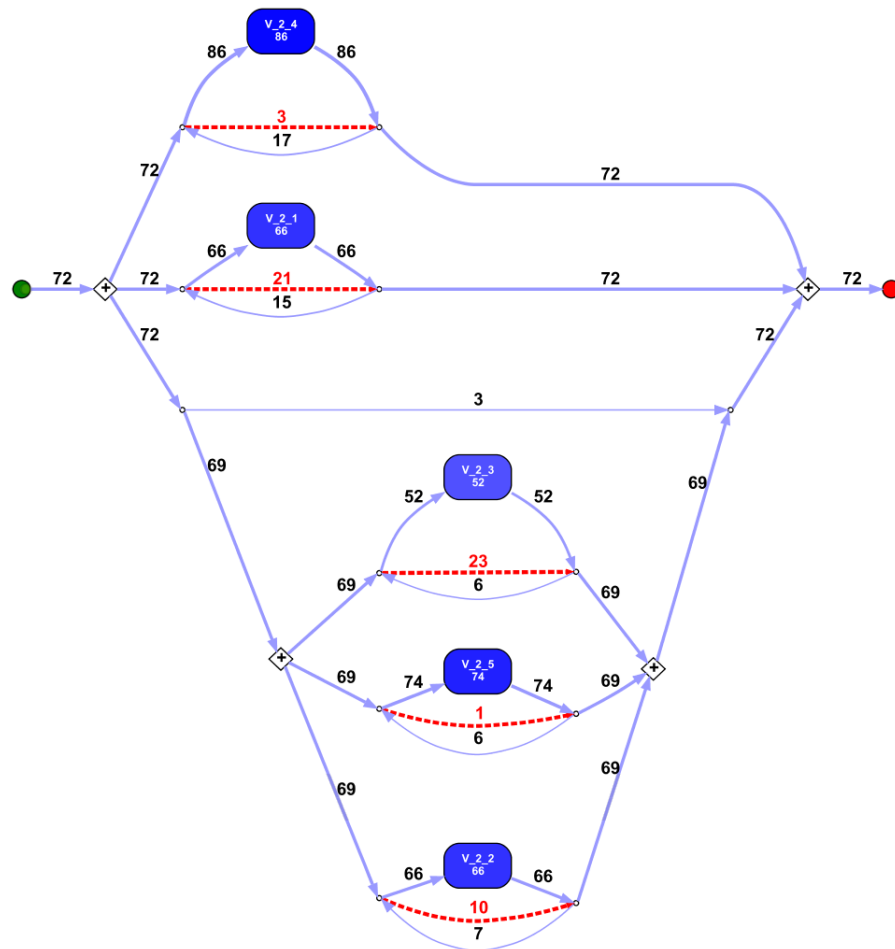


Abbildung 3: IvM-Analyse der Lehrveranstaltungen/Prüfungen des zweiten Lehrplansemesters für eine Kohorte (Rückblick aus dem sechsten Fachsemester)

² Rot gestrichelte Linien kennzeichnen durch Abstraktion begründete Abweichungen (Deviations) zwischen dem blau dargestellten Modell und den aufgezeichneten Log-Daten. Bei den in Abb. 3 dargestellten Abweichungen handelt es sich um Model-Moves, d.h. um die Kennzeichnung von Aktivitäten im Modell, die nicht von allen Prozessinstanzen ausgeführt wurden. Außerdem sind Log-Moves möglich (hier nicht dargestellte, selbst-referenzierende, rot gestrichelte Linien), die ausdrücken, dass einige Aktivitäten, die in den Logs einzelner Prozessinstanzen registriert sind, im Modell nicht wiedergegeben sind (siehe Leemans et al., 2014).

³ Minimale Einschränkung: Es werden nur die Studenten berücksichtigt, die bis zum sechsten Fachsemester mindestens eine Prüfung des zweiten Lehrplansemesters abgelegt haben.

5 Vorgehensmodell zur Studienverlaufsanalyse mit Process-Mining

Das Vorgehensmodell zur Studienverlaufsanalyse mit Process-Mining besteht aus den vier Phasen *Datenbereitstellung*, *Datenaufbereitung*, *Anwendung der Process-Mining-Verfahren* und *Ergebnisauswertung*. Diese vier Phasen werden im Folgenden weiter detailliert.

5.1 Datenbereitstellung

Studienverlaufsdaten werden im hier betrachteten Vorgehensmodell durch Ereignisse repräsentiert. Ereignisse können z.B. das Bestehen oder das Nicht-Bestehen einer Prüfung sein. Erfahrungsgemäß ist für aussagekräftige Analyseergebnisse ein Ereignis-Log mit mindestens eine vierstellige Anzahl an Ereignissen erforderlich.

Das hier beschriebene Vorgehensmodell setzt voraus, dass die auszuwertenden Studienverlaufsdaten in zwei Excel-Tabellen nach dem im Folgenden angegebenen Schema vorliegen. Die erste Excel-Tabelle wird für Fuzzy-Mining- und Inductive-Visual-Mining-Analysen benötigt, die zweite für Bubble-Chart-Analysen. Das für den Inductive Visual Miner in ProM benötigte XES-Eingabeformat erzeugen wir komfortabel mithilfe des Disco-Tools.

Tabelle 1: Die erste Excel-Tabelle muss zu jedem Ereignis die folgenden Spalten enthalten:

- **Matrikelnummer**
Die Matrikelnummer muss zu jedem Ereignis aufgeführt werden.
- **Veranstaltung**
Die Veranstaltungs(Prüfungs)-Bezeichnung muss zu jedem Ereignis aufgeführt werden. Außerdem sollte das Lehrplansemester, in dem diese Prüfung idealerweise zu absolvieren ist, angegeben werden.
- **Semester**
Das Semester ist das individuelle Fachsemester, in dem der Student das Ereignis erzeugt hat. Das konkrete Jahr, in dem er eine Prüfung absolviert, ist dagegen nicht relevant.
- **Ergebnis**
Das Ergebnis sagt aus, ob ein Student eine Prüfung erfolgreich (BE) oder nicht erfolgreich (NB) absolviert hat.

Zusätzlich zu den notwendigen Spalten ist es sinnvoll, weitere Spalten einzuführen, um weitere Analysen zu ermöglichen.

- **Versuch**
Versuch gibt an, ob ein Student eine Prüfung zum ersten Mal schreibt oder ob er sie bereits häufiger versucht hat.
- **Exmatrikulation**
Unter Exmatrikulation wird festgehalten, ob der Student, der ein Ereignis erzeugt, noch in der Hochschule immatrikuliert oder bereits exmatrikuliert ist.
- **Exmatrikulationsgrund**
Um weiterführende Informationen zu der Exmatrikulation zu erhalten, kann es sinnvoll sein, auch den Exmatrikulationsgrund aufzuführen. Hier bietet es sich an, Kategorien zu bilden.

Tabelle 2: Die zweite Excel-Tabelle (siehe Abb. 4) nimmt aggregierte Daten für die Bubble-Chart-Analysen auf. Basierend auf der ersten Excel-Tabelle wird pro Prüfung und Fachsemester eine Zeile mit folgenden Spalten generiert: Veranstaltungs-/Prüfungsbezeichnung, das betrachtete Fachsemester, die Anzahl Studenten, die die Prüfung in diesem Fachsemester bestanden bzw. nicht bestanden haben sowie die Durchfallrate (in %).

5.2 Datenaufbereitung

Vor der Ausführung der einzelnen Process-Mining-Verfahren kann es notwendig sein, die in den Excel-Tabellen vorhandenen Daten um inhaltliche und formale Fehler zu bereinigen. Falls zum Beispiel Ereignisse im 0. Semester aufgeführt sind, so sind diese zu löschen bzw. zu korrigieren. Ähnliches gilt, wenn in einem Studiengang Studenten z.B. nach unterschiedlichen Prüfungsordnungen (alt/neu) studieren. In diesem Fall muss, um die Vergleichbarkeit sicherzustellen, eine Version herausgefiltert werden.

5.3 Anwendung der Process-Mining-Verfahren

5.3.1 Bubble-Chart-Analyse

Für die Bubble-Chart-Analyse sind die Daten wie in Abb. 4 gezeigt aufzubereiten (siehe Kapitel 5.1):

Veranstaltung	Semester	Anzahl	Durchfallrate	Durchgefallene	Bestandene
V_1_1	1	66	0,00	0	66
V_1_2	1	20	45,00	9	11
V_1_3	1	75	8,00	6	69
V_1_4	1	44	59,09	26	18
V_1_5	1	73	26,03	19	54
V_1_6	1	61	86,89	53	8
...

Abbildung 4: Input-Daten (Ausschnitt) für die Bubble-Chart-Analyse einer Kohorte (Rückblick aus dem sechsten Fachsemester)

Der RapidMiner kann die Tabelle aus Abb. 4 mithilfe der Funktion *Read Excel* importieren. Unter dem Reiter *Advanced Charts* ist es dann möglich, das Blasendiagramm entsprechend zu konfigurieren. Der *Domain Dimension* wird das Fachsemester der Prüfungskandidaten zugeordnet, der *Color Dimension* die Durchfallrate, der *Size Dimension* die Anzahl und der *Nominal Axis* die Veranstaltungs-/Prüfungsbezeichnung. Zusätzlich kann bei Bedarf unter dem Reiter *Size Dimension* ein *Lower Filter* z.B. von 5 gewählt werden, um das Ergebnis visuell auf den gewünschten Abstraktionsgrad einzustellen. Prüfungen die von weniger als fünf Studenten belegt wurden, werden dann im Kontext dieser Analyse nicht berücksichtigt. Nach Einstellung des Farbschemas liefert das Programm dann für die ersten sechs Fachsemester (X-Achse) eines beispielhaften Curriculum (sechs Lehrplansemester auf der Y-Achse) das in Abb. 5 wiedergegebene Ergebnis.



Abbildung 5: Bubble-Chart-Analyse der ersten sechs Fachsemester einer Kohorte für ein beispielhaftes Curriculum (sechs Lehrplansemester)

5.3.2 Fuzzy Mining

Für die Fuzzy-Mining-Analyse mit Disco müssen zunächst die Daten der ersten Excel-Tabelle in das Programm eingelesen werden. Disco bietet dazu eine komfortable Import-Funktion, mit der auch die Semantik der einzelnen Spalten (Activity, Time, Case⁴) festgelegt werden kann. Eine Activity definieren wir in unserem Fall als Kombination aus Veranstaltung/Prüfung, Ergebnis und Versuch. Die Matrikelnummer bezeichnet den einzelnen Fall (Case) und das Fachsemester den Zeitpunkt (Time) der Prüfung. Auf Basis dieser Festlegung kann ein initiales Modell über alle Activities und Cases gebildet werden. Da wir hier einzelne Veranstaltung/Prüfung untersuchen, werden anschließend über einen Filter die Activities selektiert, die sich ausschließlich auf die ausgewählte Veranstaltung/Prüfung beziehen.

5.3.3 Inductive Visual Mining

Auch für die Inductive-Visual-Mining-Analyse mit ProM sind einige vorbereitende Schritte erforderlich. Zunächst werden die Daten aus der ersten Excel-Tabelle in Disco importiert und dort die Semantik der einzelnen Spalten festgelegt. Als Activity definieren wir hier die Veranstaltung/Prüfung⁵, als Time das Fachsemester und als

⁴ Activity, Time und Case sind vom Tool Disco gegebene Bezeichnungen.

⁵ Bei IvM bezieht sich eine Activity auf eine konkrete Veranstaltung/Prüfung, während eine Activity beim Fuzzy Mining eine Kombination aus Veranstaltung/Prüfung, Ergebnis und Versuch darstellt.

Case-Bezeichner die Matrikelnummer. Die so parametrisierten Daten werden nun als ProM-kompatible XES-Datei exportiert und in ProM geladen, wo anschließend mit dem Inductive Visual Miner das gewünschte Prozessmodell erzeugt wird. Dabei lässt sich z.B. die Wiedergabe von Abweichungen (Deviations) oder der Abstraktionsgrad der Darstellung (Prozentsatz der angezeigten Kanten und Knoten) einstellen.

5.4 Ergebnisauswertung

5.4.1 Bubble-Chart-Analyse

Die Bubble-Chart-Analyse schafft Transparenz, wenn es um die Betrachtung von Studienverläufen als Ganzes geht. Abb. 5 deutet z.B. darauf hin, dass sich sehr viele Studenten nicht an das vorgegebene Regel-Curriculum gehalten haben. Es ist zu erkennen, dass vier von sechs Prüfungen des ersten Semesters eine Durchfallquote von mindestens 25% und die Mehrzahl der Prüfungen der ersten drei Semester eine Durchfallquote von 25% oder weniger haben. Auffällig ist die Durchfallquote der Prüfung „V1_6“ im ersten Lehrplansemester (zwischen 75% und 100%). Außerdem fällt im ersten Semester auf, dass die Prüfungen zu „V_1_1“, „V_1_3“ und „V_1_5“ besonders viele Teilnehmer haben, während die Prüfungen zu „V_1_2“ und „V_1_4“ nur sehr wenig gewählt werden.

Die Ursachen für die ermittelten Ergebnisse können vielfältig sein und müssen im konkreten Einzelfall genauer untersucht werden. Möglicherweise Ansatzpunkte bei schlechten Prüfungsergebnissen: Prüfungen zu schwer, fachliche Grundlagen nicht ausreichend gelegt, falsche Erwartungshaltung bzgl. der Inhalte des Studium, Unwissen der Studenten über notwendige Voraussetzungen, Überlastung der Studenten durch parallele Veranstaltungen etc. Auf Basis identifizierter Ursachen können dann ggf. gezielte Gegenmaßnahmen ergriffen werden.

5.4.2 Fuzzy-Mining

Da die Veranstaltung V_1_6 in der Bubble-Chart-Analyse eine hohe Durchfallquote hatte, bietet es sich an, diese mithilfe von Fuzzy-Mining näher zu analysieren. In Abb. 2 ist zu sehen, wie viele Studenten die Prüfung im Erstversuch bestanden haben und wie viele durchgefallen sind. Ferner ist zu sehen, wie viele Studenten die Prüfung bereits ein zweites Mal absolviert haben und bei wie vielen Studenten die Prüfung noch offen ist. So ist erkennbar, dass 29 (=23+8-2) Studenten die Prüfung noch ablegen müssen, während 40 (=14+26-2+2) Studenten die Prüfung bereits bestanden haben. Es wird deutlich, dass etwa zwei Drittel der Studenten, die die Prüfung im Erstversuch nicht bestehen, diese im Zweitversuch bestehen.

Ein weiterer Ansatz zur Fuzzy-Mining-Analyse von Prüfungsergebnissen wäre z.B. eine vergleichende Selektion von Regelstudent bzw. Nicht-Regelstudenten, um so spezifische Merkmale des Prüfungsverhaltens beider Gruppen zu erkennen. Weiter ließe sich diese Analyse z.B. auch auf Abbrecher fokussieren, um ggf. spezielle Handlungsbedarfe (z.B. die Notwendigkeit zusätzliche Fördermaßnahmen) zu erkennen.

5.4.3 Inductive Visual Mining

Abb. 2 zeigt die Prüfungen des zweiten Lehrplansemesters. Im Beobachtungszeitraum von sechs Semestern haben insgesamt 72 Studenten mindestens eine Prüfung des zweiten Fachsemesters abgelegt. Von diesen 72 Studenten haben 3 weder V_2_2 noch V_2_3 und V_2_5 absolviert. Dies impliziert, dass Studenten diese Prüfung auf einen späteren Zeitpunkt des Studiums schieben, was den restlichen Studienablauf stören könnte. Ferner wird deutlich dass 26 (23+3) die Veranstaltung V_2_3 nicht absolviert haben. Daraus wird deutlich, dass die Prüfung zu V_2_3 die schwerste dieses Lehrplansemesters zu sein scheint. Vor allem wenn man Sie mit der Prüfung V_2_4 vergleicht, die 69 Studenten absolviert haben.

So wie beim Fuzzy-Miner wäre es auch hier möglich, Vergleiche zwischen Abbrechern, Regelstudenten und den übrigen aktiven Studenten durchzuführen, um Muster und Unterschiede beim Prüfungsverhalten zu erkennen.

6 Zusammenfassung und Fazit

Process Mining zeichnet sich durch die Möglichkeit zur Erstellung von Prozessmodellen auf Basis umfangreicher Event-Logs aus. Im hier betrachteten Anwendungsfall wurde gezeigt, wie Process-Mining-Techniken zur Analysen von Studienverläufen auf unterschiedlichen Abstraktionsebenen und aus verschiedenen Perspektiven genutzt werden kann.

Die Bubble-Chart-Analyse erlaubt eine Untersuchung des Studiums als Ganzes. Sie kann z.B. ermitteln, welche Veranstaltungen die Studenten in welchen Fachsemestern absolvieren und wie viele der Studenten bestanden bzw. nicht bestanden haben. Hieraus können bei Bedarf z.B. Maßnahmen zur Veränderung des Curriculums oder des Anforderungsniveaus von Lehrveranstaltungen und Prüfungen ableiten werden.

Die Inductive-Visual-Mining-Analyse ist bei der Analyse einzelner Lehrplansemester hilfreich. Mit ihr kann z.B. festgestellt werden, wie viele Studenten einzelne Prüfungen absolviert haben und welche Prüfungen sie gerne auf spätere Semester verschieben. Durch Vergleich mehrerer Kohorten können außerdem zeitliche Veränderungen in der Prüfungsbelegung erkannt und möglichen Ursachen zugeordnet werden (z.B. Änderungen am Veranstaltungs- und Prüfungsstoff, Änderungen der Prüfungsordnung etc.).

Mit der Fuzzy-Mining-Analyse ist es möglich, einzelne Veranstaltungen bzw. Prüfungen im Detail zu untersuchen. Dies betrifft z.B. das Belegungsverhalten der Studenten für einzelne Prüfungen, d.h. wie viele von ihnen nach dem Erstversuch einen Zweitversuch machen wie viele die Veranstaltung nach einem Fehlversuch erstmal auf einen späteren Zeitpunkt schieben. Weiter lässt sich durch eine Analyse der Verbesserungsversuche erkennen, ob Studenten eher zufrieden oder unzufrieden mit ihren Note sind.

Die verwendeten Process-Mining-Techniken ermöglichen ermöglichen ein tiefergehendes Verständnis des Prüfungsverhalten der Studenten. Dadurch können Probleme innerhalb des Curriculums entdeckt und geeignete Maßnahmen bei Bedarf und nach weiterführenden Analysen eingeleitet werden.

7 Literatur

- Emond B, Buffett S (2015) Analyzing Student Inquiry Data Using Process Discovery and Sequence Classification. Proc. 8th International Conference on Educational Data Mining, Madrid, Spain
- Günther C, van der Aalst W (2007) Fuzzy Mining – Adaptive Process Simplification Based on Multi-Perspective Metrics. Proc. 5th Int. Conf. on Business Process Management, Brisbane, Australia
- Günther C, Rozinat A (2012) Disco: Discover Your Processes. Proc. 10th Int. Conf. On Business Process Management (BPM 2012), Tallinn, Estonia
- Leemans S, Fahland D, van der Aalst W (2013). Discovering block-structured process models from event logs containing infrequent behaviour. Proc. 11th Int. Conf. on Business Process Management, Beijing, China
- Leemans S, Fahland D, van der Aalst W (2014) Exploring Processes and Deviations. Proc. 12th Int. Conf. on Business Process Management (BPM 2014), Eindhoven, The Netherlands
- Mierswa I (2009) Open Source Data Mining mit RapidMiner. Künstliche Intelligenz 2/09
- Mukala P, Buijs J, van der Aalst W (2015) Exploring Students' Learning Behaviour in MOOCs using Process Mining Techniques. BPM Center Report BPM-15-10, Eindhoven University of Technology, The Netherlands
- Process Mining Group (2016) XES (eXtensible Event Stream). <http://www.processmining.org/logs/xes> Zugriff 27.11.2016
- Rischawe R, Buck-Emden R (2015) Process-Mining in der Assekuranz. HMD Praxis der Wirtschaftsinformatik, 52(3)
- Trčka N, Pechenizkiy M, van der Aalst W (2010) Process Mining from Educational Data. In: Romero C et al. (Eds.) Handbook of Educational Data Mining, CRC Press
- van der Aalst W. (2016) Process Mining. Springer Heidelberg New York
- van der Aalst W, Weijters A (2004) Process Mining: A Research Agenda. Computers in Industry 53(3)
- van Dongen B, de Medeiros A, Verbeek H, Weijters A, van der Aalst W (2005) The ProM Framework: A New Era in Process Mining Tool Support. Proc. 26th Int. Conf. On Applications and Theory of Petri Nets (ICATPN 2005), Miami, FL